

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
12. April 2001 (12.04.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/26262 A2

(51) Internationale Patentklassifikation⁷: **H04H 1/00**

(21) Internationales Aktenzeichen: PCT/EP00/09771

(22) Internationales Anmeldedatum:
5. Oktober 2000 (05.10.2000)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
199 47 877.5 5. Oktober 1999 (05.10.1999) DE

(71) Anmelder (*für alle Bestimmungsstaaten mit Ausnahme von US*): **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Leonrodstrasse 54, 80636 München (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (*nur für US*): **NEUBAUER, Christian** [DE/DE]; Effeltricher Strasse 24, 90411 Nürnberg (DE). **HERRE, Jürgen** [DE/DE]; Am Eichengarten 11, 91054 Buckenhof (DE). **BRANDENBURG, Karlheinz** [DE/DE]; Haagstrasse 32, 91054 Erlangen (DE). **ALLAMANCHE, Eric** [DE/DE]; Sulzbacherstrasse 41, 90489 Nürnberg (DE).

(74) Anwälte: **SCHOPPE, Fritz usw.**; Schoppe, Zimmermann & Stöckeler, Postfach 71 08 67, 81458 München (DE).

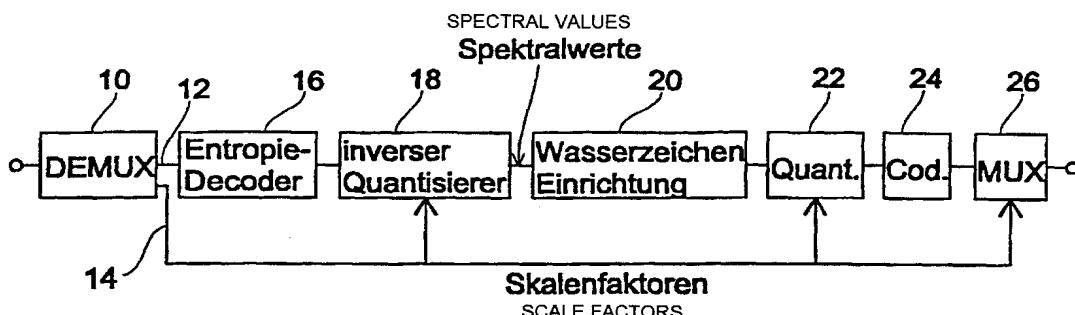
(81) Bestimmungsstaat (national): US.

(84) Bestimmungsstaaten (regional): europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD AND DEVICE FOR INTRODUCING INFORMATION INTO A DATA STREAM AND A METHOD FOR ENCODING AN AUDIO SIGNAL

(54) Bezeichnung: VERFAHREN UND VORRICHTUNG ZUM EINBRINGEN VON INFORMATIONEN IN EINEN DATENSTROM SOWIE VERFAHREN UND VORRICHTUNG ZUM CODIEREN EINES AUDIOSIGNALS



16... ENTROPIC DECODES
18... INVERSE QUANTIFIER
20... WATERMARK DEVICE

WO 01/26262 A2

(57) Abstract: The invention relates to a method for introducing information into a data stream containing data related to spectral values which represent a short-term spectrum of an audio signal. According to said method, the data stream is first processed (10, 16, 18), in order to obtain the spectral values for the short-term spectrum of the audio signal. In addition, the information to be introduced is subjected to a spread sequence (20) in order to obtain an expanded information signal which leads to the creation of a spectral representation of the expanded information signal (20). This representation is then weighted using a determined psychoacoustic noise energy which can be masked (20), in order to generate a weighted information signal, in which the energy level of the introduced information is substantially equal to or lies below the psychoacoustic masking threshold. The weighted information signal and the spectral values for the short-term spectrum are subsequently totalled (20) and then re-processed (22, 24, 26) in order to obtain a processed data stream which comprises both the audio information and the information to be introduced. In order for the information to be introduced without having to pass into the time domain, the block raster which underlies the short-term spectrum is not infringed, so that the introduction of a watermark does not lead to a tandem encoding effect.

[Fortsetzung auf der nächsten Seite]

**Veröffentlicht:**

— Ohne internationales Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) Zusammenfassung: Ein erfindungsgemäßes Verfahren zum Einbringen von Informationen in einen Datenstrom, der Daten Über Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen, führt zuerst eine Verarbeitung (10, 16, 18) des Datenstroms durch, um die Spektralwerte des Kurzzeitspektrums des Audiosignals zu erhalten. Außerdem werden die einzubringenden Informationen mit einer Spreizsequenz beaufschlagt (20), um ein gespreiztes Informationssignal zu erhalten, woraufhin eine spektrale Darstellung des gespreizten Informationssignals erzeugt wird (20), die dann mit einer ermittelten psychoakustisch maskierbaren Störenergie gewichtet wird (20), um ein gewichtetes Informationssignal zu erzeugen, bei dem die Energie der eingebrachten Informationen im Wesentlichen gleich oder unterhalb der psychoakustischen Maskierungsschwelle liegt. Das gewichtete Informationssignal und die Spektralwerte des Kurzzeitspektrums des Audiosignals werden dann summiert (20) und anschließend wieder verarbeitet (22, 24, 26), um einen verarbeiteten Datenstrom zu erhalten, der sowohl die Audioinformationen als auch die einzubringenden Informationen umfaßt. Dadurch, daß die einzubringenden Informationen in den Datenstrom eingebbracht werden, ohne daß in den Zeitbereich Übergegangen werden muß, wird die Blockrasterung, die dem Kurzzeitspektrum zugrunde liegt, nicht angetastet, so daß das Einbringen eines Wasserzeichens zu keinen Tandem-Codiereffekten führt.

- 1 -

**Verfahren und Vorrichtung zum Einbringen von Informationen
in einen Datenstrom sowie Verfahren und Vorrichtung zum
Codieren eines Audiosignals**

Beschreibung

Die vorliegende Erfindung bezieht sich allgemein auf Audiosignale und insbesondere auf das Einbringen von Informationen in einen Datenstrom, der Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen. Insbesondere auf dem Gebiet der Urheberrechtsschutzes für Audiosignale dient die vorliegende Erfindung dazu, beispielsweise Urheberrechtsinformationen in ein Audiosignal möglichst unhörbar einzubringen.

Mit zunehmender Verbreitung des Internets hat auch die Musikpiraterie drastisch zugenommen. An vielen Stellen im Internet können Musikstücke bzw. allgemein Audiosignale heruntergeladen werden. In den allerwenigsten Fällen werden hierbei Urheberrechte beachtet. Insbesondere wird sehr selten die Erlaubnis des Urhebers eingeholt, sein Werk zur Verfügung zu stellen. Noch seltener werden Gebühren an den Urheber bezahlt, die der Preis für ein rechtmäßiges Kopieren sind. Darüberhinaus findet ein unkontrolliertes Kopieren von Werken statt, was in den allermeisten Fällen ebenfalls ohne Berücksichtigung von Urheberrechten geschieht.

Wenn Musikstücke über das Internet von einem Provider für Musikstücke rechtmäßig erworben werden, erzeugt der Provider üblicherweise einen Header, in dem Copyright-Informationen sowie beispielsweise eine Kundennummer eingebracht sind, wobei die Kundennummer eindeutig auf den aktuell vorliegenden Käufer hinweist. Es ist ferner bekannt, Kopiererlaubnisinformationen in diesen Header einzufügen, welche die verschiedensten Arten von Kopierrechten signalisieren, wie z.B. daß das Kopieren des aktuellen Stücks vollständig untersagt ist, daß das Kopieren des aktuellen Stücks nur ein

- 2 -

einziges Mal erlaubt ist, daß das Kopieren des aktuellen Stücks völlig frei ist, etc.

Der Kunde verfügt über einen Decodierer, der den Header einliest und unter Beachtung der erlaubten Handlungen beispielsweise nur eine einzige Kopie zuläßt und weitere Kopien verweigert.

Dieses Konzept zur Beachtung der Urheberrechte funktioniert jedoch nur für Kunden, die sich legal verhalten.

Illegalen Kunden haben üblicherweise ein wesentliches Potential an Kreativität, um mit einem Header versehene Musikstücke zu "knacken". Hier zeigt sich bereits der Nachteil der beschriebenen Vorgehensweise zum Schutz von Urheberrechten. Ein solcher Header kann einfach entfernt werden. Alternativ könnte ein illegaler Benutzer auch einzelne Einträge in dem Header modifizieren, um beispielsweise aus dem Eintrag "Kopieren untersagt" einen Eintrag "Kopieren völlig frei" zu machen. Denkbar ist auch der Fall, daß ein illegaler Kunde seine eigene Kundennummer aus dem Header entfernt und dann das Musikstück auf seiner oder einer anderen Homepage im Internet anbietet. Ab diesem Moment ist es nicht mehr möglich, den illegalen Kunden zu ermitteln, da er seine Kundennummer entfernt hat. Versuche, solche Verletzungen des Urheberrechts zu unterbinden, werden daher zwangsläufig ins Leere laufen, da die Kopierinformationen aus dem Musikstück entfernt worden sind bzw. modifiziert worden sind, und da der illegale Kunde, der dies tat, nicht mehr ermittelt werden kann, um ihn zur Verantwortung zu ziehen. Wäre stattdessen eine sichere Einbringung von Informationen in das Audiosignal vorhanden, so könnten staatliche Behörden, die Urheberrechtsverletzungen verfolgen, verdächtige Musikstücke im Internet ermitteln und beispielsweise die Benutzeridentifikation solcher illegalen Stücke feststellen, um den illegalen Benutzern das Handwerk zu legen.

Aus der WO 97/33391 ist ein Codierverfahren zur Einbringung

eines nicht hörbaren Datensignals in ein Audiosignal bekannt. Dabei wird das Audiosignal, in das das nicht hörbare Datensignal eingebracht werden soll, in den Frequenzbereich umgewandelt, um mittels eines psychoakustischen Modells die Maskierungsschwelle des Audiosignals zu bestimmen. Das Datensignal, das in das Audiosignal eingebracht werden soll, wird mit einem Pseudorauschsignal multipliziert, um ein frequenzmäßig gespreiztes Datensignal zu schaffen. Das frequenzmäßig gespreizte Datensignal wird dann mit der psychoakustischen Maskierungsschwelle gewichtet, derart, daß die Energie des frequenzmäßig gespreizten Datensignals immer unterhalb der Maskierungsschwelle liegt. Schließlich wird das gewichtete Datensignal dem Audiosignal überlagert, wodurch ein Audiosignal erzeugt wird, in das das Datensignal unhörbar eingebracht ist. Das Datensignal kann zum einen dazu verwendet werden, die Reichweite eines Senders zu ermitteln. Alternativ kann das Datensignal zur Kennzeichnung von Audiosignalen verwendet werden, um eventuelle Raubkopien ohne weiteres zu identifizieren, da jeder Tonträger beispielsweise in Form einer CompactDisc ab Werk mit einer individuellen Kennung versehen wird. Weitere beschriebene Anwendungsmöglichkeiten des Datensignals bestehen im Fernsteuern von Audiogeräten in Analogie zum "VPS"-Verfahren beim Fernsehen.

Dieses Verfahren liefert bereits eine hohe Sicherheit gegenüber Musikpiraten, da sie zum einen unter Umständen gar nicht wissen, daß das Musikstück, das sie gerade vervielfältigen, gekennzeichnet ist. Darüberhinaus ist es nahezu unmöglich, ohne einen autorisierten Decodierer das Datensignal, das unhörbar in dem Audiosignal vorhanden ist, zu extrahieren.

Audiosignale liegen, wenn sie von einer CompactDisc stammen, als 16-Bit-PCM-Abtastwerte vor. Ein Musikpirat könnte beispielsweise die Abtastrate bzw. die Pegel oder Phasen der Abtastwerte manipulieren, um das Datensignal unlesbar, d. h. undecodierbar zu machen, wodurch die Urheberrechtsinformationen ebenfalls aus dem Audiosignal entfernt wären. Dies

wird jedoch nicht ohne signifikante Qualitätseinbußen möglich sein. Solchermaßen in Audiosignale eingebrachte Daten können daher auch in Analogie zu Banknoten als "Wasserzeichen" bezeichnet werden.

Das in der WO 97/33391 beschriebene Verfahren zur Einbringung eines nicht hörbaren Datensignals in ein Audiosignal arbeitet unter Verwendung der Audioabtastwerte, die als Zeitbereichs-Abtastwerte vorliegen. Dies macht es erforderlich, daß Audiostücke, d. h. Musikstücke, Hörspiele und ähnliches, als Folge von zeitlichen Abtastwerten vorliegen müssen, um mit einem Wasserzeichen versehen zu werden. Dies hat den Nachteil, daß dieses Verfahren nicht für bereits komprimierte Datenströme, die beispielsweise einer Verarbeitung nach einem der MPEG-Verfahren unterzogen worden sind, eingesetzt werden kann. Dies bedeutet, daß ein Anbieter von Musikstücken, der die Musikstücke vor der Auslieferung an den Kunden mit einem Wasserzeichen versehen möchte, die Musikstücke als Folge von PCM-Abtastwerten speichern muß. Dies führt dazu, daß ein Anbieter für Musikstücke eine überaus große Speicherkapazität zur Verfügung haben muß. Es wäre jedoch wünschenswert, die äußerst effektiven Audiokomprimierungsverfahren bereits zum Speichern der Audiodaten bei dem Anbieter einzusetzen.

Selbstverständlich könnte ein Anbieter für Audiodaten der oben beschriebenen Art einfach hergehen, sämtliche Musikstücke beispielsweise unter Verwendung des Standards MPEG-2 AAC 13818-7 komprimieren und dann, bevor das Audiostück mit einem Wasserzeichen versehen werden soll, wieder vollständig dekomprimieren, um wieder eine Folge von Audio-Abtastwerten ("Samples") zu haben, die dann in eine bekannte Vorrichtung zum Einbringen eines nicht hörbaren Datensignals eingespeist werden, um ein Wasserzeichen einzubringen. Dies bringt jedoch einen wesentlichen Aufwand dahingehend mit sich, daß vor dem Einbringen der Informationen in das Audiosignal eine vollständige Dekomprimierung bzw. Decodierung erforderlich ist. Eine solche Decodierung kostet Zeit und Geld. Wesent-

lich gravierender ist jedoch die Tatsache, daß bei einer solchen Vorgehensweise Tandem-Codiereffekte auftreten.

Ein weiterer Nachteil dieser Vorgehensweise besteht darin, daß aufgrund der Tatsache, daß das Wasserzeichen in die PCM-Daten eingebracht wird, keine Sicherheit darüber gegeben ist, ob das Wasserzeichen nach einer Audio-Komprimierung noch vorhanden ist. Wenn mit Wasserzeichen versehene PCM-Daten mit relativ niedriger Bitrate codiert werden, führt der Codierer beim Quantisieren aufgrund der relativ niedrigen Bitrate viel Quantisierungsrauschen ein, das im extremen Fall dazu führen wird, daß kein Wasserzeichen mehr decodiert werden kann. Problematisch ist also, daß bei dieser Vorgehensweise die Bitrate des Audiocodierers, der die mit Wasserzeichen versehenen PCM-Daten codiert, nicht von vorneherein bekannt ist, weshalb keine sichere Steuerung des Verhältnisses zwischen Wasserzeichenenergie und Störenergie aufgrund des Quantisierungsrauschens möglich ist.

Es ist bekannt, daß Audio-Coderverfahren nach einem der MPEG-Standards keine verlustlosen Coderverfahren, sondern verlustbehaftete Coderverfahren sind. Biteinsparungen im Vergleich zur direkten Übertragung von Audio-Abtastwerten im Zeitbereich werden zu einem Großteil dadurch erreicht, daß psychoakustische Maskierungseffekte ausgenutzt werden. Insbesondere wird für einen Block von beispielsweise 2048 Audio-Abtastwerten die psychoakustische Maskierungsschwelle als Funktion der Frequenz ermittelt, woraufhin nach einer Zeit-Frequenz-Transformation der Audio-Abtastwerte die Quantisierung der Spektralwerte, die das Kurzzeitspektrum enthält, unter Berücksichtigung dieser psychoakustischen Maskierungsschwelle durchgeführt wird. Anders ausgedrückt wird die Quantisierer-Schrittweite so gesteuert, daß die durch das Quantisieren eingefügte Störenergie kleiner oder gleich der psychoakustischen Maskierungsschwelle ist. In Bereichen des Audiosignals, in denen das Verdeckungsmaß, d. h. das Verhältnis der Audiosignalenergie und der psychoakustischen Maskierungsschwelle, sehr klein ist, wie z. B. in sehr

rauschhaften Bereichen des Audiosignals, müssen die Spektralwerte lediglich grob quantisiert werden, ohne daß es nach einem anschließenden Decodieren zu hörbaren Störungen kommt. In anderen Bereichen, in denen das Audiosignal sehr tonal ist, muß feiner quantisiert werden, derart, daß eine relativ kleine Störennergie aufgrund des Quantisierens entsteht, da das Verdeckungsmaß hier sehr groß ist.

Aus dem Vorstehenden wird deutlich, daß aufgrund der Quantisierungsvorgehensweise Informationen des ursprünglichen Audiosignals verloren gehen. Dies spielt noch keine Rolle, wenn das quantisierte Audiosignal wieder decodiert wird, da die Störennergie aufgrund des Quantisierens so verteilt wurde, daß sie unter der psychoakustischen Maskierungsschwelle bleibt und somit, wenn ein ideales psychoakustisches Modell verwendet wurde, unhörbar sein wird. Diese Betrachtungen gelten jedoch immer nur für ein bestimmtes Kurzzeitspektrum bzw. für einen Block von z. B. 2.048 aufeinanderfolgenden Audio-Abtastwerten. Nach der Decodierung enthält der Block von Audio-Abtastwerten jedoch keine Informationen mehr darüber, wie die Blockbildung durchgeführt wurde. Wenn die bekannte Vorrichtung zum Einbringen von Informationen verwendet wird, die in den allermeisten Fällen eine bestimmte Verzögerung im Vergleich zu einem Audio-Codierer, der keine Informationen einbringt, hat, kann nicht davon ausgegangen werden, daß rein zufällig die gleiche Blockeinteilung stattfindet. Stattdessen werden die Blockeinteilung, die Kurzzeit-Spektrum-Bildung und die Quantisierung in einem völlig anderen Blockraster stattfinden. Eine erneute Decodierung wird dann üblicherweise zu deutlich hörbaren Störungen führen, da sie sich nicht auf das gleiche Kurzzeitspektrum sondern auf unterschiedliche Kurzzeitspektren bezieht. Dieses Auftreten von hörbaren Störungen durch zwei Codierer/Decodierer-Stufen aufgrund ihrer unterschiedlichen Einteilung des Stroms von Audio-Abtastwerten in Blöcke wird als Tandem-Codiereffekt bezeichnet.

Es sei hier darauf hingewiesen, daß im allgemeinen Fall

durch das Einbringen des nicht-hörbaren Datensignals Stör-energie in das Audiosignal eingeführt wird, das ohnehin schon Störerenergie aufgrund des nicht unendlich feinen Quantisierungsverfahrens hat. Das Einbringen des nicht-hörbaren Datensignals führt damit tendenziell zu einer Verschlechterung der Audioqualität, es sei denn, daß besondere Vorkehrungen unternommen werden. In diesem Zusammenhang ist eine zusätzliche Einführung von Störerenergie aufgrund der Tandem-Codiereffekte, die vorstehend ausgeführt wurden, umso weniger wünschenswert, da dieser Qualitätsverlust einfach systembedingt ohne Nutzen auftritt, während kleine Qualitätsverschlechterungen aufgrund des Wasserzeichens eher in Kauf genommen werden, da das Wasserzeichen einen Nutzen mit sich bringt. Tandem-Codiereffekte bringen jedoch nur Störungen, aber überhaupt keinen Nutzen mit sich.

Die Aufgabe der vorliegenden Erfindung besteht darin, ein Konzept zu schaffen, das es ermöglicht, Audiostücke mit einem Wasserzeichen zu versehen, während die Auswirkungen des Wasserzeichens auf die Audioqualität möglichst gering sein sollen.

Diese Aufgabe wird durch ein Verfahren zum Einbringen von Informationen in einen Datenstrom nach Patentanspruch 1, durch ein Verfahren zum Codieren eines Audiosignals nach Patentanspruch 11 oder 12, durch eine Vorrichtung zum Einbringen von Informationen nach Patentanspruch 13 und durch eine Vorrichtung zum Codieren eines Audiosignals nach Patentanspruch 15 oder Patentanspruch 16 gelöst.

Der vorliegenden Erfindung liegt die Erkenntnis zugrunde, daß davon abgegangen werden muß, vor dem Einbringen des Wasserzeichens eine vollständige Decodierung durchzuführen. Stattdessen wird erfindungsgemäß ein Datenstrom, der Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen, lediglich teilweise "entpackt", bis die Spektralwerte vorliegen. Die Entpackung ist jedoch keine vollständige Decodierung sondern lediglich eine teilweise

Decodierung, bei der die Informationen über die Blockbildung bzw. das im ursprünglichen Codierer eingesetzte Blockraster nicht angetastet werden.

Dies wird dadurch erreicht, daß das erfindungsgemäße Verfahren mit Spektralwerten und nicht mit zeitlichen Abtastwerten (Samples) durchgeführt wird. Die Informationen, die in das Audiosignal eingebracht werden sollen, werden im Sinne eines Spread-Spectrum-Modulation mit einer Spreizsequenz beaufschlagt, um ein gespreiztes Informationssignal zu erhalten. Anschließend wird eine spektrale Darstellung des gespreizten Informationssignals beispielsweise durch eine Filterbank, eine FFT, einer MDCT oder ähnliches erzeugt, um ein spektrales gespreiztes Informationssignal zu erhalten. Nun wird eine psychoakustisch maskierbare Störung als Funktion der Frequenz für das Kurzzeitspektrum des Audiosignals ermittelt, um dann das spektrale gespreizte Informationssignal unter Verwendung der ermittelten Störennergie zu gewichten, so daß ein gewichtetes Informationssignal erzeugt werden kann, dessen Energie im wesentlichen gleich oder unterhalb der psychoakustischen Maskierungsschwelle liegt. Daran anschließend wird das gewichtete Informationssignal mit den Spektralwerten des Kurzzeitspektrums des Audiosignals summiert, um Summen-Spektralwerte zu erhalten, die das Kurzzeitspektrum des Audiosignals und zusätzlich die eingebrachten Informationen umfassen. Schließlich werden die Summen-Spektralwerte wieder verarbeitet, um einen verarbeiteten Datenstrom zu erhalten, der die Daten über die Spektralwerte des Kurzzeitspektrums des Audiosignals und die einzubringenden Informationen umfaßt. Im Falle eines MPEG-AAC-Codierers wird das Verarbeiten der Summen-Spektralwerte wieder das Quantisieren und Entropie-Codieren beispielsweise unter Verwendung eines Huffman-Codes zum Gegenstand haben.

Es sei darauf hingewiesen, daß dadurch nicht die Blockrasterung, die der ursprüngliche Codierer festgelegt hat, der den Datenstrom erzeugt hat, angetastet wird. Dadurch entstehen keine Tandem-Effekte, die zu einem Verlust der Audioqualität

führen würden. Darüberhinaus wird es bevorzugt, daß bei der nach dem Gewichten erfolgenden Verarbeitung, die ein Quantisieren umfaßt, die gleiche bzw. die gleichen Quantisierungsschrittweite(n) wie im ursprünglichen Bitstrom verwendet wird bzw. werden, was den Vorteil mit sich bringt, daß die sehr rechenaufwendigen Iterationsschleifen des Quantisierers nicht erneut berechnet werden müssen. Ferner treten keine Tandem-Codier-Effekte auf, die ansonsten unvermeidbar wären, da im Falle einer erneuten Berechnung mehr oder weniger stark abweichende Quantisiererschrittweiten auftreten könnten.

Das erfindungsgemäße Einbringen eines Wasserzeichens direkt in einen Datenstrom ermöglicht beispielsweise das Einbringen einer Kundennummer während der Auslieferung der Musik an einen Kunden, da das Verfahren auf modernen Personalcomputern in mehrfacher Echtzeit ablaufen kann, da u. a. zusätzlich auf die aufwendige Frequenz-Zeit-Transformation verzichtet werden kann, die bei einer vollständigen Decodierung erforderlich sein würde.

Ein weiterer Vorteil der vorliegenden Erfindung besteht darin, daß der Musikanbieter, d. h. der Musikprovider, nicht mehr die PCM-Abtastwerte speichern muß, sondern vorcodierte Datenströme speichern kann, was im Speicherplatz durchaus einen Faktor in der Größenordnung von 12 mit sich bringen kann, und daß der Provider trotzdem kundenspezifische Wasserzeichen einbringen kann, ohne daß zusätzlich Tandem-Codiereffekte auftreten würden, die einen Audioqualitätsverlust zur Folge hätten.

Das erfindungsgemäße Verfahren kann einfach implementiert werden, da nur eine zusätzliche Zeit/Frequenz-Transformation des gespreizten Informationssignals erforderlich ist. Ein weiterer wesentlicher Vorteil besteht darin, daß das erfindungsgemäße Verfahren eine gute Interoperabilität besitzt, d. h. daß Standard-Datenströme verarbeitet werden können, und daß für Wasserzeichen gemäß den bekannten Verfahren und

für Wasserzeichen gemäß dem erfindungsgemäßen Verfahren der gleiche Wasserzeichendecodierer verwendet werden kann. Schließlich besteht ein weiterer Vorteil darin, daß ein Audiodcodierer das Wasserzeichen nicht mehr auslöschen kann, da eine genaue Steuerung des Verhältnisses zwischen Quantisierungsrauschen und Wasserzeichenenergie besteht.

Es sei darauf hingewiesen, daß es selbstverständlich möglich ist, das Wasserzeichen unbefugt wieder zu entfernen, wenn der mit einem Wasserzeichen versehene Datenstrom decodiert wird und dann wieder codiert wird, jedoch nun mit niedriger Bitrate. In diesem Fall wird die durch den Quantisierer eingeführte Störenergie die Wasserzeichenenergie übersteigen, so daß dann kein Wasserzeichen mehr aus dem Audiosignal extrahiert werden kann. Dies ist jedoch unproblematisch, da die Audioqualität des Audiosignals aufgrund des hohen Quantisierungsrauschen derart stark abgenommen hat, daß ein solches schlechtes Audiosignal auch nicht mehr geschützt werden muß. Ist nämlich in einem Audiosignal das Wasserzeichen zerstört, so ist auch seine Qualität zerstört.

Die psychoakustisch maskierbare Störenergie kann auf verschiedene Arten und Weisen ermittelt werden. Eine erste Option besteht darin, zum Ermitteln der psychoakustisch maskierbaren Störenergie ein psychoakustisches Modell einzusetzen, das aus dem Kurzzeitspektrum die psychoakustische Maskierungsschwelle als Funktion der Frequenz erzeugt. Es existiert eine Vielzahl von psychoakustischen Modellen, wobei hier die psychoakustischen Modelle besonders von Vorteil sind, die ohnehin mit Spektralwerten des Kurzzeitspektrums arbeiten, da diese Spektralwerte direkt aufgrund des teilweisen Entpackens des Datenstroms vorliegen. Alternativ können jedoch auch psychoakustische Modelle zum Einsatz kommen, die für Zeitbereichsdaten ausgestaltet sind, wobei hier im Gegensatz zu der oben beschriebenen Option eine Frequenz-Zeit-Transformation erforderlich sein würde. Obwohl die Möglichkeit des Berechnens eines psychoakustischen Modells, um die psychoakustische Maskierungsschwelle des Kurzzeitspek-

trums zu erhalten, relativ rechenzeitaufwendig ist, liefert jedoch bereits diese Möglichkeit den entscheidenden Vorteil, daß keine Tandem-Codiereffekte erzeugt werden, da die Blockrasterung nicht angetastet wird.

Eine weitere, im Rechenzeitaufwand günstigere Option zum Ermitteln der psychoakustisch maskierbaren Störennergie besteht darin, daß der Datenstrom derart erzeugt worden ist, daß er neben den Spektralwerten und den üblichen Seiteninformationen auch für jedes Kurzzeitspektrum die psychoakustische Maskierungsschwelle als Funktion der Frequenz enthält. Ein Ermitteln der psychoakustisch maskierbaren Störennergie funktioniert dann einfach durch Extrahieren der im Datenstrom übertragenen psychoakustischen Maskierungsschwelle. Bei dieser Möglichkeit und der vorher beschriebenen Möglichkeit, bei der das psychoakustische Maskierungsmodell berechnet wird, ist die psychoakustisch maskierbare Störennergie die psychoakustische Maskierungsschwelle selbst. Nachteilig an dem Verfahren des Übertragens der psychoakustischen Maskierungsschwelle im Datenstrom ist die Tatsache, daß ein Spezial-Audiocodierer benötigt wird, da bei üblichen Audio-Codierern die psychoakustische Maskierungsschwelle nicht übertragen wird, sondern lediglich die Spektralwerte und die entsprechenden Skalenfaktoren. In geschlossenen Systemen ist jedoch Kompatibilität zu Standard-Datenströmen nicht erforderlich. Hier kann diese Option daher mit einfacherem Aufwand und rechenzeitgünstig implementiert werden.

Eine weitere Möglichkeit besteht darin, einen Spezial-Audio-codierer vorzusehen, dessen Quantisierer immer so arbeitet, daß das Quantisierungsrauschen um einen vorbestimmten Betrag geringer als die psychoakustische Maskierungsschwelle ist. Dies bedeutet, daß der Codierer so ausgelegt ist, daß sein Quantisierer etwas feiner quantisiert, als er eigentlich müßte, derart, daß zusätzliche Störennergie hinzugefügt werden kann, ohne daß eine Störung hörbar wird. Diese zusätzliche Störennergie kann dann beim Einbringen von Informationen in den Datenstrom "aufgebraucht" werden, um die Informa-

tionen einzubringen. Im Falle eines optimalen psychoakustischen Modells führt diese Möglichkeit zu einem Datenstrom mit eingebrachtem Wasserzeichen, der überhaupt keine Qualitätsverschlechterung erlitten hat. Nachteilig an diesem Verfahren ist ebenso wie beim direkten Übertragen der psychoakustischen Maskierungsschwelle die Tatsache, daß dieses Verfahren nicht mit üblichen Codierern kompatibel ist.

Eine weitere Möglichkeit zum Ermitteln der psychoakustisch maskierbaren Störennergie besteht darin, die tatsächlich durch das Quantisieren des Codierers, der den Datenstrom erzeugt hat, eingebrachte Störennergie zu ermitteln und daraus die psychoakustisch maskierbare Störennergie, die die Informationen beim Gewichten erhalten werden, abzuleiten. Diese Option geht davon aus, daß der Codierer so quantisiert hat, daß die Störennergie unter der psychoakustischen Maskierungsschwelle oder lediglich knapp darüber gelegen hat. Dieses Verfahren kommt ebenfalls wie das als erste Möglichkeit beschriebene Verfahren mit den Standard-Bitströmen aus, da lediglich die Spektralwerte und die Skalenfaktoren, die beide im Datenstrom vorhanden sind, benötigt werden, um die psychoakustisch maskierbare Störennergie zu erhalten. Aus den Skalenfaktoren kann die Schrittgröße des Quantisierers, der dem entsprechenden Skalenfaktor zugeordnet ist, ermittelt werden, um damit die in einem Skalenfaktorbau eingebrauchte Störennergie zu errechnen, die typischerweise gleich der psychoakustischen Maskierungsschwelle sein wird oder darunter liegen wird. Die beim Gewichten verwendete psychoakustisch maskierbare Störennergie für die eingebrachten Informationen kann gleich der Quantisierungsstörennergie sein, sie kann aber auch einen Faktor zwischen größer als Null und kleiner als Eins haben, wobei ein Faktor näher bei Null zu weniger hörbaren Störungen aufgrund des Wasserzeichens führen wird, aber beim Extrahieren problematischer sein könnte als ein Faktor näher bei Eins.

Bevorzugte Ausführungsbeispiele der vorliegenden Erfindung werden nachfolgend bezugnehmend auf die beiliegenden Zeich-

nungen detailliert erläutert. Es zeigen:

Fig. 1 ein Blockdiagramm einer erfindungsgemäßen Vorrichtung zum Einbringen von Informationen in einen Datenstrom;

Fig. 2 ein detailliertes Blockdiagramm der Wasserzeichen-einrichtung von Fig. 1;

Fig. 3a eine schematische Darstellung eines Verfahrens zum Ermitteln der maskierbaren Störenergie unter Verwendung eines psychoakustischen Modell;

Fig. 3b eine schematische Darstellung eines Verfahrens zum Ermitteln der maskierbaren Störenergie, wenn die psychoakustische Maskierungsschwelle im Datenstrom übertragen wird;

Fig. 3c eine schematische Darstellung eines Verfahrens zum Ermitteln der maskierbaren Störenergie, wenn die Störenergie unter Kenntnis der Spektralwerte und der Skalenfaktoren geschätzt wird;

Fig. 3d eine schematische Darstellung eines Verfahrens zum Ermitteln der psychoakustisch maskierbaren Störenergie, wenn im Datenstrom Energie für das Wasserzeichen freigehalten wird; und

Fig. 4 ein Blockdiagramm eines erfindungsgemäßen Audiocodierers, der entweder die psychoakustische Maskierungsschwelle in den Datenstrom schreibt, oder den vorbestimmten Betrag für das in Fig. 3d beschriebene Verfahren in den Datenstrom schreibt und dessen Quantisierer entsprechend gesteuert ist.

Bevor detaillierter auf die einzelnen Figuren eingegangen wird, wird der systemtheoretische Hintergrund der vorliegenden Erfindung kurz beleuchtet. Generell darf das Einbringen

der Informationen in das Audiosignal zu keiner bzw. nur einer sehr schwer hörbaren Qualitätsverschlechterung des Audiosignals führen. Um festzustellen, wieviel Energie das Signal, das die einzubringenden Informationen darstellt, haben darf, wird unter Verwendung eines psychoakustischen Modells die Maskierungsschwelle des Audiosignals fortlaufend berechnet. Die frequenzselektive Berechnung der Maskierungsschwelle unter Verwendung beispielsweise der kritischen Bänder sowie eine Vielzahl weiterer psychoakustischer Modelle sind in der Technik bekannt. Beispielhaft wird auf den Standard MPEG2-AAC (ISO/IEC 13818-7) verwiesen.

Das psychoakustische Modell führt zu einer Maskierungsschwelle für ein Kurzzeitspektrum des Audiosignals. Üblicherweise wird die Maskierungsschwelle über der Frequenz variieren. Per Definition wird angenommen, daß ein in das Audiosignal eingebrachtes Signal dann unhörbar sein wird, wenn die Energie dieses Signals unterhalb der Maskierungsschwelle ist. Die Maskierungsschwelle hängt stark von der Zusammensetzung des Audiosignals ab. Rauschhafte Signale haben eine höhere Maskierungsschwelle als sehr tonale Signale. Die Energie des Signals, das in das Audiosignal eingebracht wird, variiert daher stark über der Zeit. Üblicherweise wird zum Decodieren der in ein Audiosignal eingebrachten Informationen ein bestimmtes Signal/Rausch-Verhältnis benötigt. Dabei kann es vorkommen, daß bei sehr tonalen Audiosignalabschnitten die Energie des zusätzlich eingebrachten Signals derart gering wird, daß das Signal/Rausch-Verhältnis zum sicheren Decodieren nicht mehr ausreicht. Ein Decodierer kann daher in solchen Bereichen einzelne Bits nicht mehr korrekt decodieren. Systemtheoretisch gesehen kann daher das Einbringen von Informationen in ein Audiosignal in Abhängigkeit von der psychoakustischen Maskierungsschwelle als das Übertragen eines Datensignals über einen Kanal mit stark variierender Störenergie betrachtet werden, wobei das Audiosignal, d. h. das Musiksignal, als Störsignal aufgefaßt wird.

Fig. 1 zeigt ein Blockdiagramm einer erfindungsgemäßen Vor-

richtung bzw. eines erfindungsgemäßen Verfahrens zum Einbringen von Informationen in einen Datenstrom, der Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen. Der Datenstrom, der am Eingang eines Datenstrom-Demultiplexers 10 anliegt, wird, wenn er gemäß dem erwähnten MPEG-AAC-Standard verarbeitet ist, zunächst allgemein in Spektralwerte auf einer Leitung 12 und Seiteninformationen auf einer Leitung 14 zerlegt, wobei von den Seiteninformationen hier die Skalenfaktoren speziell genannt seien. Die Spektralwerte, die hinter dem Demultiplexer 10 noch Entropie-codiert sind, werden dann einem Entropie-Decodierer 16 zugeführt und dann einem inversen Quantisierer 18, der unter Verwendung der quantisierten Spektralwerte und der dazu gehörigen Skalenfaktoren, die über die Leitung 14 dem inversen Quantisierer 18 zu Verfügung gestellt werden, die Spektralwerte des Audiosignals erzeugt, die das Kurzzeitspektrum desselben darstellen. Die Spektralwerte werden dann in eine Wasserzeicheneinrichtung 20 eingespeist, die Summen-Spektralwerte erzeugt, die das Kurzzeitspektrum des Audiosignals und darüber hinaus die einzubringenden Informationen umfassen. Diese Summen-Spektralwerte werden dann wieder in einen Quantisierer 22 gespeist und in einem daran anschließenden Entropie-Codierer 24 Entropie-codiert, um schließlich einem Datenstrom-Multiplexer 26 zugeführt zu werden, der auch wieder die nötigen Seiteninformationen, wie z. B. die Skalenfaktoren, erhält. Am Ausgang des Multiplexers 26 liegt dann ein verarbeiteter Datenstrom vor, der sich vom Datenstrom am Eingang des Demultiplexers 10 darin unterscheidet, daß er nun ein Wasserzeichen hat, d. h. daß in ihn Informationen eingebracht worden sind.

Bevor näher auf Fig. 2 eingegangen wird, die eine detailliertere Darstellung der Wasserzeicheneinrichtung 20 aufweist, sei zum Verständnis auf einen MPEG-2 AAC-Audiocodierer eingegangen, wie er beispielsweise im Anhang B des Standard ISO/IEC 13818-7:1997(E) als informativer Teil beschrieben ist. Einem solchen Codierer liegt grundsätzlich die Idee zugrunde, das Quantisierungsrauschen unter die

sogenannte psychoakustische Maskierungsschwelle zu bringen, d. h. zu verstecken. Zur Transformation der Audio-Abtastwerte in den Frequenzbereich, d. h. zum Erzeugen der spektralen Darstellung des Audiosignals wird eine Analysefilterbank eingesetzt, die als kritisch-unterabgetastete DCT (DCT = diskrete Cosinustransformation) realisiert ist, und die einen Überlappungsgrad von 50% hat. Ihr Zweck besteht darin, eine spektrale Darstellung des Eingangssignals zu schaffen, das schließlich quantisiert und codiert wird. Zusammen mit einer entsprechenden Filterbank im Decodierer entsteht somit ein Synthese/Analyse-System.

Das psychoakustische Modell, das in solchen Codierern verwendet wird, basiert auf dem psychoakustischen Phänomen der Maskierung. Sowohl Frequenzbereichsmaskierungseffekte als auch Zeitbereichsmaskierungseffekte können dabei modelliert werden. Das psychoakustische Modell liefert einen Schätzwert für "Rausch"-Energie, die dem ursprünglichen Audiosignal hinzugefügt werden kann, ohne daß hörbare Störungen auftreten. Diese maximal zulässige Energie wird als psychoakustische Maskierungsschwelle bezeichnet.

Der Quantisierer 22 sowie der Codierer 24 in Fig. 1 werden im nachfolgenden beschrieben. Typischerweise wird mehr als eine Spektrallinie mit derselben Quantisierer-Schrittgröße quantisiert. Daher werden mehrere benachbarte Spektrallinien in sogenannte Skalenfaktorbänder gruppiert. Der Quantisierer optimiert die Quantisiererschrittgröße für jedes Skalenfaktorband. Die Quantisiererschrittgröße wird so bestimmt, daß der Quantisierungsfehler unter oder gleich der berechneten psychoakustischen Maskierungsschwelle ist, um sicherzustellen, daß das Quantisierungsrauschen unhörbar ist. Es ist zu sehen, daß zwei Begrenzungen berücksichtigt werden müssen, zwischen denen ein Kompromiß gefunden werden muß. Einerseits sollte der Bitverbrauch so niedrig als möglich gehalten werden, um hohe Kompressionsverhältnisse, d. h. einen hohen Codiergewinn, zu erreichen. Andererseits muß sichergestellt werden, daß das Quantisierungsrauschen unter der psychoaku-

stischen Maskierungsschwelle ist, damit im codierten und wieder decodierten Audiosignal keine Störungen hörbar sind. Typischerweise wird dieses Optimierungsverfahren in einer iterativen Schleife berechnet. Das Resultat dieser Schleife ist eine Quantisiererschrittgröße, die mit einem Skalenfaktor für ein Skalenfaktorband eindeutig korrespondiert. Anders ausgedrückt werden die Spektralwerte eines Skalenfaktorbandes mit einer Quantisiererschrittgröße quantisiert, die dem für das Skalenfaktorband maßgeblichen Skalenfaktor eindeutig zugeordnet ist. Das heißt, daß zwei verschiedene Skalenfaktoren auch zwei verschiedene Quantisiererschrittgrößen zur Folge haben können.

Der Bitstrom wird durch einen Bitstrommultiplexer zusammengesetzt, der im wesentlichen Formatierungsaufgaben erfüllt. Der Datenstrom, der im Falle eines Binärsystems ein Bitstrom ist, enthält somit die quantisierten und codierten Spektralwerte oder Spektralkoeffizienten sowie die Skalenfaktoren und weitere Seiteninformationen, die im erwähnten MPEG-AAC-Standard detailliert dargestellt und erläutert sind.

Fig. 2 zeigt ein detaillierteres Blockdiagramm der Wasserzeicheneinrichtung 20 von Fig. 1. An einer Quelle 30 für Informationseinheiten werden Informationseinheiten, vorzugsweise in Form von Bits, einer Einrichtung 32 zum Spreizen zugeführt. Die Einrichtung 32 zum Spreizen basiert grundsätzlich auf einer Spread-Spektrum-Modulation, die insbesondere unter Verwendung einer Pseudo-Noise-Spreizsequenz bei einer Korrelation im Wasserzeichenextraktor günstig ist. Die Informationen werden Bit für Bit mit der Spreizsequenz beaufschlagt. Das Beaufschlagen findet vorzugsweise so statt, daß für ein Informationsbit mit einem logischen Pegel von +1 die Spreizsequenz unverändert am Ausgang der Einrichtung 32 erzeugt wird, während für ein Informationsbit mit einem logischen Pegel von 0, was beispielsweise einem Spannungspiegel von -1 entsprechen kann, die umgekehrte Spreizsequenz am Ausgang der Einrichtung 32 erzeugt wird. Damit entsteht ein "Zeitsignal" am Ausgang der Einrichtung 32, das die ge-

spreizten Informationen aus der Quelle 30 für Informationen enthält. Dieses gespreizte Informationssignal wird dann durch eine Einrichtung 34 zum Transformieren in seine spektrale Darstellung überführt, die ein FFT-Algorithmus, eine MDCT, etc., aber auch eine Filterbank sein kann. Die spektrale Darstellung des gespreizten Informationssignals wird in einer Einrichtung 36 gewichtet, um dann mit den Spektralwerten in einer Einrichtung 38 summiert zu werden, derart, daß am Ausgang der Einrichtung 38 die Summen-Spektralwerte anliegen, die dann bezugnehmend auf Fig. 1 quantisiert (22) und codiert (24) werden können, um dem Bitstrom-Multiplexer 26 zugeführt zu werden. Die Wasserzeicheneinrichtung 20 enthält ferner eine Einrichtung 40 zum Ermitteln der maskierbaren Störenergie für das Kurzzeitspektrum, das durch die Spektralwerte gegeben ist.

Es sei darauf hingewiesen, daß die Einrichtung 34 zum Transformieren des gespreizten Informationssignals vorzugsweise eine Spektraltransformation durchführt, die der dem Datenstrom am Eingang des Demultiplexers 10 (Fig. 1) zugrunde liegenden Transformation entspricht. Das heißt, daß die Einrichtung 34 zum Transformieren vorzugsweise dieselbe modifizierte diskrete Cosinustransformation durchführt, die ursprünglich zum Erzeugen des nicht-verarbeiteten Datenstroms verwendet wurde. Dies ist ohne weiteres möglich, da Informationen, wie z. B. Fenstertyp, Fensterform, Fensterlänge usw., als Seiteninformationen im Bitstrom übertragen werden. Diese Verknüpfung ist durch die in Fig. 2 gestrichelt dargestellte Linie vom Bitstrom-Demultiplexer 10 (Fig. 1) ange deutet.

Wie es bereits bezugnehmend auf Fig. 1 ausgeführt worden ist, werden die Summen-Spektralwerte nach der Addition im Summierer 38 wieder einer Quantisierung und Codierung unterzogen. Hier stellt sich die Frage, wie das Quantisiererintervall, d. h. die Quantisiererschrittgröße, auf die bereits eingegangen worden ist, bestimmt werden soll, d. h. ob die Iterationen erneut durchgeführt werden müssen, oder ob dar-

auf verzichtet werden kann. Aufgrund der Tatsache, daß die Wasserzeichenenergie im Vergleich zur Audiosignalenergie üblicherweise sehr klein ist, können vorzugsweise dieselben Skalenfaktoren wie im ursprünglichen Bitstrom eingesetzt werden. Dies ist in Fig. 1 durch die Verbindungsleitung 14 vom Demultiplexer 10 zum Multiplexer 26 dargestellt. Das heißt, daß das Quantisieren durch den Quantisierer 22 wesentlich einfacher durchgeführt werden kann, da es nicht mehr notwendig ist (aber dennoch möglich ist), die Iterationsschleifen auszuführen, um einen optimalen Kompromiß zwischen Bitrate und Quantisiererschrittgröße zu bestimmen. Stattdessen werden vorzugsweise einfach die bereits bekannten Skalenfaktoren verwendet.

Im nachfolgenden wird auf verschiedene Möglichkeiten eingegangen, um die durch das Kurzzeitspektrum maskierbare Störenergie zu ermitteln, die beim Gewichten der spektralen Darstellung des gespreizten Informationssignals benötigt wird. Hierzu existieren verschiedene Möglichkeiten, die nachfolgend bezugnehmend auf die Fig. 3a-3d erläutert sind.

In Fig. 3a wird ein psychoakustisches Modell eingesetzt, um unter Verwendung der Spektralwerte des Audiosignals die psychoakustische Maskierungsschwelle des entsprechenden Kurzzeitspektrums zu errechnen. Aufgrund der Tatsache, daß psychoakustische Modelle in der Literatur und dem erwähnten Standard beschrieben sind, sei hier lediglich erwähnt, daß vorzugsweise psychoakustische Modelle verwendet werden können, die ohnehin mit Spektraldaten arbeiten bzw. eine Zeit/Frequenz-Transformation beinhalten. In diesem Fall ist das psychoakustische Modell zum ursprünglichen psychoakustischen Modell, das einem jedem Codierer zugrunde liegt, darin vereinfacht, daß dasselbe gleich mit Spektralwerten "gefüttert" werden kann, so daß überhaupt keine Frequenz/Zeit-Transformation im psychoakustischen Modell erforderlich ist. Das psychoakustische Modell schließlich wird die psychoakustische Maskierungsschwelle für das Kurzzeitspektrum ausgeben, derart, daß im Block 36 (Fig. 2) das

Spektrum des gespreizten Informationssignals so geformt werden kann, daß es in jedem Skalenfaktorband eine Energie hat, die gleich der psychoakustischen Maskierungsschwelle bzw. unterhalb der psychoakustischen Maskierungsschwelle in diesem Skalenfaktorband ist. Es sei darauf hingewiesen, daß die psychoakustische Maskierungsschwelle eine Energie ist, wobei es angestrebt wird, daß die spektrale Darstellung des Informationssignals möglichst gleich der psychoakustischen Maskierungsschwelle ist, um die Informationen in das Audiosignal durch möglichst viel Energie einzubringen, um in einem Extraktor des Wasserzeichens möglichst gute Korrelationsspitzen zu erhalten.

Die in Fig. 3a gezeigte erste Möglichkeit hat den Vorteil, daß die psychoakustische Maskierungsschwelle sehr genau berechnet werden kann, und daß dieses Verfahren mit üblichen Datenströmen vollständig kompatibel ist. Nachteilig darin ist jedoch die Tatsache, daß die Berechnung eines psychoakustischen Modells üblicherweise relativ zeitaufwendig sein kann, so daß gesagt werden kann, daß diese Möglichkeit zwar sehr genau und interoperabel ist, jedoch relativ viel Zeit benötigt.

Eine weitere Möglichkeit, die in Fig. 3b gezeigt ist, um die psychoakustisch maskierbare Störennergie zu erhalten, besteht darin, daß im Codierer, der den Datenstrom am Eingang des Demultiplexers 10 (Fig. 1) erzeugt hat, die psychoakustische Maskierungsschwelle für jedes Kurzzeitspektrum in den Bitstrom geschrieben wird, derart, daß die erfundungsgemäße Vorrichtung zum Einbringen von Informationen in einen Datenstrom lediglich die psychoakustische Maskierungsschwelle für jedes Kurzzeitspektrum aus den Seiteninformationen des Datenstroms zu extrahieren braucht (40b), um die psychoakustische Maskierungsschwelle zur Einrichtung 36 zum Gewichten der spektralen Darstellung des gespreizten Informationssignals (Fig. 2) auszugeben. Diese Möglichkeit hat den Vorteil, daß sie ebenfalls sehr genau ist und darüberhinaus sehr schnell ist, da lediglich zugegriffen und nicht gerech-

net werden muß, es wird jedoch die Interoperabilität beeinträchtigt, d. h. Standardbitströme können nicht mehr nachträglich mit Wasserzeichen versehen werden, da sie keine psychoakustischen Maskierungsschwellen enthalten. Man benötigt hier somit einen erfindungsgemäßen Spezialcodierer, wie er in Fig. 4 beschrieben ist.

In Fig. 3c ist eine weitere Möglichkeit zum Ermitteln der psychoakustisch maskierbaren Störennergie gezeigt. Hier wird die psychoakustisch maskierbare Störennergie unter Verwendung der Spektralwerte und der Skalenfaktoren berechnet (40c). Es wird davon ausgegangen, daß der ursprüngliche Codierer, der den Datenstrom, in den das Wasserzeichen eingebracht werden soll, erzeugt hat, die durch die Quantisierung eingeführte Störennergie bereits so gewählt hat, daß sie unter der psychoakustischen Maskierungsschwelle bzw. auf der psychoakustischen Maskierungsschwelle liegt. Dieses Verfahren ist zwar etwas ungenauer als das direkte Berechnen der psychoakustischen Maskierungsschwelle, ist jedoch im Vergleich zum direkten Berechnen der psychoakustischen Maskierungsschwelle sehr schnell und hält gleichzeitig die Interoperabilität aufrecht, d. h. arbeitet auch mit Standard-Bitströmen zusammen.

Im nachfolgenden wird darauf eingegangen, wieso diese dritte Möglichkeit etwas ungenauer ist. Es existieren verschiedene Codiereransätze, die sich beispielsweise in den verwendeten Quantisiererimplementationen unterscheiden. Wie es bereits ausgeführt worden ist, darf ein Quantisierer die vorgeschriebene Bitrate nicht überschreiten. Andererseits soll er die psychoakustische Maskierungsschwelle einhalten. So kann der Fall auftreten, daß ein Quantisierer die zur Verfügung stehende Bitrate gar nicht benötigt, da beispielsweise eine hohe Bitrate vorliegt, oder wenn ein Musikstück zu codieren ist, bei dem der CodiergeWINN sehr hoch ist, wie es beispielsweise bei tonalen Stücken der Fall ist. Bestimmte Quantisierer arbeiten hier so, daß sie feiner als nötig quantisieren und somit wesentlich weniger Störennergie durch-

Quantisieren in das Audiosignal einführen, als sie dürften. Es ist daher einsichtig, daß die erfindungsgemäße Vorrichtung, wie sie in Fig. 3c beschrieben ist, davon ausgeht, daß die psychoakustische Maskierungsschwelle wesentlich geringer ist als sie tatsächlich sein dürfte, was schließlich dazu führt, daß die spektrale Darstellung des gespreizten Informationssignals nach dem Gewichten wesentlich weniger Energie hat als sie haben dürfte, wodurch nicht die ganze verfügbare Energie, die das Wasserzeichen haben darf, ausgenutzt wird. Dies würde jedoch nicht der Fall sein, wenn ein Quantisierer eingesetzt wird, der immer die maximal zulässige Störennergie durch das Quantisieren einführt und eventuell verbleibende Bits nicht beschreibt bzw. mit irgendwelchen bei einer Decodierung nicht berücksichtigten Werten füllt. In diesem Fall wäre die in Fig. 3c dargestellte Option genauso genau wie die beiden ersten Möglichkeiten. Im Falle des variablen Quantisierers entsteht jedoch auch eine variable Bitrate. In diesem Fall könnte die Wasserzeicheneinrichtung auch dazu verwendet werden, die Bitrate konstant zu machen, durch Auffüllen von Bits, die das Wasserzeichen darstellen, so daß die konstante Bitrate gleich der höchsten Bitrate des ursprünglichen Datenstroms mit variabler Bitrate ist.

Im nachfolgenden wird darauf eingegangen, wie unter Verwendung der Spektralwerte und der Skalenfaktoren und darüberhinaus der Charakteristik des Quantisierers die Störennergie berechnet wird, die durch Quantisieren in ein Skalenfaktorband eingeführt worden ist. Hierbei gilt folgende Gleichung für die Energie F_{xi} des Quantisierungsfehlers für einen Spektralwert x_i :

$$|F_{xi}|^2 = (q^{2\alpha}/12\alpha^2) \cdot x_i^{2(1-\alpha)}$$

Es sei darauf hingewiesen, daß diese Gleichung für ungleichmäßige Quantisierer gilt, wie sie beispielsweise bei dem Standard MPEG-AAC vorgesehen sind. Für gleichmäßige Quantisierer würde der zweite Term einfach wegfallen, wenn für $\alpha =$

1 gesetzt wird.

Der in der Gleichung auftretende Faktor q hängt mit der Quantisiererschrittgröße QS folgendermaßen zusammen:

$$q = 2^{QS/4}$$

Der Faktor α lautet für den MPEG-AAC-Quantisierer 3/4.

Die Energie des Quantisierungsfehlers in einem Skalenfaktorband ist dann die Summe der $|Fxi|^2$ in einem Skalenfaktorband. Diese Energie muß, damit sie nicht hörbar ist, kleiner oder gleich der psychoakustischen Maskierungsschwelle in diesem Skalenfaktorband sein. Es sei darauf hingewiesen, daß die psychoakustische Maskierungsschwelle in einem Skalenfaktorband konstant ist, jedoch für unterschiedliche Skalenfaktorbänder unterschiedliche Werte einnimmt. Für die Energie des Quantisierungsfehlers x_{min} ergibt sich folgender Wert:

$$x_{min} = \sum_i [(2^{3/8} \cdot QS) / (27/4) \cdot x_i^{1/2}]$$

Der Index i soll anzeigen, daß immer über die Spektralwerte in einem Skalenfaktorband summiert werden muß, da die psychoakustische Maskierungsschwelle üblicherweise als Energie für dieses Skalenfaktorband gegeben ist.

Es sei darauf hingewiesen, daß in den Seiteninformationen des Datenstroms nicht direkt die Quantisiererschrittgrößen für die einzelnen Skalenfaktoren gegeben sind, daß jedoch gemäß Vereinbarung, wie sie im AAC-Standard aufgeführt ist, die Quantisiererschrittgröße, die jedem Skalenfaktor zugeordnet ist, eindeutig abgeleitet werden kann. Darüberhinaus muß die Charakteristik des im ursprünglichen Codierer zum Erzeugen des Datenstroms verwendeten Quantisierers bekannt sein, d. h., wenn er ein ungleichmäßiger Quantisierer ist, der Kompressionsfaktor desselben, der beim AAC-Standard der

Faktor 3/4 ist.

Die Spektrallinien der spektralen Darstellung des gespreizten Informationssignals werden nun, wie es bereits ausgeführt worden ist, so gewichtet, daß sie zusammen eine Energie haben, die kleiner oder gleich der psychoakustisch maskierbaren Störennergie, und im Falle der in Fig. 3c beschriebenen Option gleich der Störennergie des Quantisierungsprozesses ist.

Wenn der Fall betrachtet wird, daß die durch die Quantisierung in Skalenfaktorband eingeführte Störennergie bereits gleich der psychoakustischen Maskierungsschwelle ist, und dann dieselbe Energie noch einmal, jedoch nun für die einzubringenden Informationen in das Audiosignal eingebracht wird, so ist zu sehen, daß die insgesamte Energie, d. h. die Störennergie aufgrund des Quantisierens und die Energie für die Informationen, die psychoakustische Maskierungsschwelle überschreiten können, was zu hörbaren Qualitätsverlusten führen kann, die jedoch aufgrund der Begrenzung der Energie der Informationen auf die psychoakustische Maskierungsschwelle klein sein werden, da die psychoakustische Maskierungsschwelle um einen Faktor größer als 1 verletzt wird. Wie es bereits ausgeführt worden ist, wird eine Wasserzeichenenergie in der Größenordnung der psychoakustischen Maskierungsschwelle dann zu Störungen führen, wenn auch das Quantisierungsrauschen bereits in der Größenordnung der psychoakustischen Maskierungsschwelle liegt. Es wird daher bevorzugt, die psychoakustisch maskierbare Störennergie, mit der gewichtet wird, so zu wählen, daß die gesamte Störennergie (Quantisierungsrauschen plus "Störennergie" der Informationen) kleiner als das 1,5-fache der psychoakustischen Maskierungsschwelle ist, wobei noch kleinere Faktoren bis nahe 1,0 möglich sind. Es sei darauf hingewiesen, daß auch kleine Faktoren sinnvoll sind, da aufgrund der Spreizung des Informationssignals bereits eine hohe Informationsredundanz eingeführt worden ist.

Anders ausgedrückt wird das Einbringen eines Wasserzeichens in ein Audiosignal, dessen psychoakustische Maskierungsschwelle bereits vollständig durch Störennergie aufgrund des Quantisierens verbraucht ist, zu einer geringen Verschlechterung der Audioqualität führen, die jedoch durch die Vorteile des Wasserzeichens leicht aufgehoben wird.

Um diese Beeinträchtigung zu überwinden, kann das in Fig. 3d gezeigte Konzept eingesetzt werden, bei dem von vornehmlich der Quantisierer im Codierer derart gesteuert wird, daß die durch die Quantisierung eingeführte Störennergie durch Einstellen der Quantisiererschrittgröße so gewählt wird, daß sie immer einen vorbestimmten Betrag unter der psychoakustischen Maskierungsschwelle bleibt. Mit anderen Worten arbeitet ein Audiocodierer für ein solches Konzept so, daß er feiner quantisiert als nötig, wodurch ein Energiepotential für die einzubringenden Informationen, d. h. für das Wasserzeichen freibleibt. Dies hat den Vorteil, daß ein Wasserzeichen vollständig ohne Qualitätsverlust eingebracht werden kann, wenn beim Ermitteln der psychoakustisch maskierbaren Störennergie (40d) die ja nun um einen vorbestimmten Betrag kleiner als die psychoakustische Maskierungsschwelle ist, in der Einrichtung 40d der vorbestimmte Betrag berücksichtigt wird, so daß die Störennergie aufgrund des Quantisierens und die Energie aufgrund der einzubringenden Informationen zusammen gleich oder kleiner als die psychoakustische Maskierungsschwelle sind. Da die gewichteten Spektralwerte des gespreizten Informationssignals mit den Spektralwerten des Audiosignals summiert werden, sind die Spektralwerte des Informationssignals nach ihrer Gewichtung gleich oder kleiner als der vorbestimmte Betrag.

Diese Option hat den Vorteil, daß ein Wasserzeichen ohne jeglichen Qualitätsverlust in einen Datenstrom eingebracht werden kann, daß jedoch einerseits die Interoperabilität leidet, und da der Quantisierer im Codierer immer um den vorbestimmten Betrag unter der psychoakustischen Maskierungsschwelle bei der Einstellung der Störennergie durch das

Quantisieren bleiben muß. Andererseits ist diese Möglichkeit in der Implementation sehr effizient, da kein psychoakustisches Modell berechnet werden muß.

Im nachfolgenden wird auf Fig. 4 eingegangen, wobei Fig. 4 zwei Möglichkeiten für einen Codierer für Audiosignale zeigt, um einen Datenstrom zu erzeugen, der erfindungsgemäß besonders für ein Einbringen von Informationen geeignet ist. Ein solcher Audiocodierer kann grundsätzlich so aufgebaut sein, wie ein bekannter Audiocodierer, derart, daß er eine Einrichtung 50 zum Erzeugen einer spektralen Darstellung des Audiosignals, einen Quantisierer 52 zum Quantisieren der spektralen Darstellung des Audiosignals, einen Entropie-Codierer 54 zum Entropie-Codieren der quantisierten Spektralwerte und schließlich einen Datenstrommultiplexer 56 umfaßt. Der Datenstrom, der von dem Datenstrom-Multiplexer 56 ausgegeben wird, erhält von einem ebenfalls bekannten psychoakustischen Modell 58 über den Datenstrommultiplexer 56 die psychoakustische Maskierungsschwelle, die im Gegensatz zu einem bekannten Audiocodierer nun in den Datenstrom geschrieben wird, derart, daß die erfindungsgemäße Vorrichtung zum Einbringen von Informationen einfach auf die psychoakustische Maskierungsschwelle in dem Datenstrom zugreifen kann. Der in Fig. 4 durch die durchgezogene Linie 60 dargestellte Codierer ist somit das Gegenstück zu der in Fig. 1 gezeigten Vorrichtung zum Einbringen von Informationen, die als Einrichtung zum Ermitteln der maskierbaren Störenergie die in Fig. 3b gezeigte Option enthält.

In Fig. 4 gestrichelt ist die Audiocodierermöglichkeit gemäß der vorliegenden Erfindung dargestellt, die zu der in Fig. 3d gezeigten Option für die Einrichtung 40 zum Ermitteln der maskierbaren Störenergie in der in Fig. 1 gezeigten erfindungsgemäßen Vorrichtung korrespondiert. Hierbei wird der Quantisierer durch einen vorbestimmten Betrag derart gesteuert, daß die durch die Quantisierung eingeführte Störenergie um den vorbestimmten Betrag unterhalb der psychoakustischen Maskierungsschwelle ist, wobei der Wert des vorbestimmten

- 27 -

Betrags über die gestrichelte Leitung 62 in den Datenstrommultiplexer 56 eingespeist wird, um in dem Datenstrom enthalten zu sein, derart, daß die erfindungsgemäße Vorrichtung zum Einbringen von Informationen auf den vorbestimmten Betrag zugreifen kann, um entsprechend gewichten zu können (Block 36 in Fig. 2).

Patentansprüche

1. Verfahren zum Einbringen von Informationen in einen Datenstrom, der Daten über Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen, mit folgenden Schritten:

Verarbeiten (10, 16, 18) des Datenstroms, um die Spektralwerte des Kurzzeitspektrums des Audiosignals zu erhalten;

Beaufschlagen (32) der Informationen mit einer Spreizsequenz, um ein gespreiztes Informationssignal zu erhalten;

Erzeugen (34) einer Spektraldarstellung des gespreizten Informationssignals, um ein spektrales gespreiztes Informationssignal zu erhalten;

Ermitteln (40a; 40b; 40c; 40d) einer psychoakustisch maskierbaren Störenergie als Funktion der Frequenz für das Kurzzeitspektrum des Audiosignals, wobei die psychoakustisch maskierbare Störenergie kleiner oder gleich der psychoakustischen Maskierungsschwelle des Kurzzeitspektrums ist;

Gewichten (36) des spektralen gespreizten Informationssignals unter Verwendung der ermittelten Störenergie, um ein gewichtetes Informationssignal zu erzeugen, bei dem die Energie der eingebrachten Informationen im wesentlichen gleich oder unterhalb der psychoakustischen Maskierungsschwelle liegt;

Summieren (38) des gewichteten Informationssignals mit den Spektralwerten des Kurzzeitspektrums des Audiosignals, um Summen-Spektralwerte zu erhalten, die das

Kurzzeitspektrum des Audiosignals und die Informationen umfassen; und

Verarbeiten (22, 24, 26) der Summen-Spektralwerte, um einen verarbeiteten Datenstrom zu erhalten, der die Daten über die Spektralwerte des Kurzzeitspektrums des Audiosignals und die einzubringenden Informationen umfaßt.

2. Verfahren nach Anspruch 1, bei dem der Datenstrom als Daten über Spektralwerte quantisierte Spektralwerte enthält, wobei der Schritt des Verarbeitens des Datenstroms folgenden Teilschritt aufweist:

inverses Quantisieren (18) der quantisierten Spektralwerte, um die Spektralwerte zu erhalten; und

bei dem der Schritt des Verarbeitens der Summen-Spektralwerte folgende Schritte aufweist:

Quantisieren (22) der Summen-Spektralwerte, um quantisierte Summen-Spektralwerte zu erhalten; und

Bilden (26) des verarbeiteten Datenstroms unter Verwendung der quantisierten Summen-Spektralwerte.

3. Verfahren nach Anspruch 2, bei dem die quantisierten Spektralwerte im Datenstrom Entropie-codiert sind, wobei der Schritt des Verarbeitens des Datenstroms folgenden Teilschritt aufweist:

Entropie-Decodieren (18) der Entropie-codierten Spektralwerte, um die quantisierten Spektralwerte zu erhalten; und

bei dem der Schritt des Verarbeitens der Summen-Spektralwerte folgenden Schritt aufweist:

Entropie-Codieren (24) der quantisierten Summen-Spektralwerte.

4. Verfahren nach einem der vorhergehenden Ansprüche, bei dem der Schritt des Ermittelns der psychoakustisch maskierbaren Störennergie folgenden Schritt aufweist:

Berechnen (40a) der psychoakustischen Maskierungsschwelle als Funktion der Frequenz unter Verwendung eines psychoakustischen Modells, das basierend auf den Spektralwerten des Audiosignals arbeitet.

5. Verfahren nach einem der Ansprüche 1 bis 3, bei dem im Datenstrom als Seiteninformationen eine beim Erzeugen des Datenstroms verwendete Maskierungsschwelle als Funktion der Frequenz für das Kurzzeitspektrum vorhanden ist, wobei der Schritt des Ermittelns folgenden Schritt aufweist:

Extrahieren (40b) der psychoakustischen Maskierungsschwelle aus dem Datenstrom, wobei die psychoakustisch maskierbare Störennergie gleich der psychoakustischen Maskierungsschwelle ist.

6. Verfahren nach einem der Ansprüche 1 bis 3, bei dem der Datenstrom ferner Seiteninformationen aufweist, die Skalenfaktoren (14) beinhalten, mit denen die Spektralwerte vor dem Quantisieren in einem Audiocodierer gruppenweise multipliziert wurden, wobei der Schritt des Verarbeitens des Datenstroms ferner folgenden Teilschritt aufweist:

Extrahieren der Skalenfaktoren aus dem Datenstrom; und

bei dem der Schritt des Ermittelns folgenden Schritt aufweist:

Berechnen der beim Quantisieren im Audiocodierer ein-

geführten Störennergie als Funktion der Frequenz unter Verwendung der Skalenfaktoren für das Kurzzeitspektrum und unter Verwendung der Spektralwerte sowie unter Kenntnis eines im Audiocodierer verwendeten Quantisierers, wobei die eingebrachte Störennergie ein Maß für die psychoakustisch maskierbare Störennergie ist, die beim Gewichten verwendet wird.

7. Verfahren nach Anspruch 6, bei dem der Datenstrom gemäß ISO/IEC 13818-7 (MPEG-2 AAC) ausgebildet ist, und bei dem der Schritt des Schätzens der Störennergie folgende Schritte aufweist:

Ermitteln eines Quantisierungsschritts für die Spektralwerte aus einem Skalenfaktorband unter Verwendung des diesem Skalenfaktorband zugeordneten Skalenfaktors;

Auswerten der folgenden Gleichung, um die durch die Quantisierung eingeführte Störennergie für das Skalenfaktorband zu erhalten,

$$x_{\min} = \sum_i [(2^{3/8} \cdot QS) / (27/4) \cdot x_i^{1/2}]$$

wobei x_i die i-te Spektrallinie in einem Skalenfaktorband darstellt, wobei QS der Quantisierungsschritt für dieses Skalenfaktorband ist, und wobei x_{\min} die durch die Quantisierung in das Skalenfaktorband eingeführte Störennergie ist;

und bei dem der Schritt des Gewichtens (36) folgenden Schritt aufweist:

Einstellen der Spektralwerte der spektralen Darstellung des gespreizten Informationssignals in dem Skalenfaktorband so, daß die Gesamtenergie der eingestellten Spektralwerte gleich der im Schritt des Auswertens erhaltenen Störennergie in diesem Skalenfaktorband ist.

8. Verfahren nach einem der Ansprüche 1 bis 3, bei dem die Spektralwerte des Datenstroms derart quantisiert sind, daß die durch die Quantisierung eingeführte Störenergie um einen vorbestimmten Betrag kleiner als die psychoakustische Maskierungsschwelle ist, und bei dem im Schritt des Ermittelns (40d) eine Energie bestimmt wird, die dem vorbestimmten Betrag entspricht; und
bei dem im Schritt des Gewichtens (36) die Spektralwerte der spektralen Darstellung des gespreizten Informationssignals derart eingestellt werden, daß sie eine Energie haben, die dem vorbestimmten Betrag entspricht.
9. Verfahren nach Anspruch 8, bei dem der Wert des vorbestimmten Betrags als Seiteninformationen in dem Datenstrom vorhanden ist, wobei im Schritt des Ermittelns (40d) der Wert für den vorbestimmten Betrag aus den Seiteninformationen des Datenstroms extrahiert wird.
10. Verfahren nach einem der vorhergehenden Ansprüche, bei dem im Schritt des Verarbeitens der Summen-Spektralwerte die gleichen Quantisierungsschrittweiten wie im ursprünglichen Datenstrom verwendet werden.
11. Verfahren zum Codieren eines Audiosignals mit folgenden Schritten:

Erzeugen (50) eines Kurzzeitspektrums des Audiosignals, das eine Mehrzahl von Spektralwerten umfaßt;

Berechnen der psychoakustischen Maskierungsschwelle des Audiosignals unter Verwendung eines psychoakustischen Modells (58);

Quantisieren (52) der Spektralwerte unter Berücksichtigung der psychoakustischen Maskierungsschwelle, so daß die durch die Quantisierung eingeführte Störenergie

gleich oder kleiner als die psychoakustische Maskierungsschwelle ist; und

Bilden (56) eines Bitstroms, der Werte enthält, die den quantisierten Spektralwerten des Kurzzeitspektrums entsprechen, und der darüberhinaus die berechnete psychoakustische Maskierungsschwelle (60) für das Kurzzeitpektrum des Audiosignals aufweist.

12. Verfahren zum Codieren eines Audiosignals, mit folgenden Schritten:

Erzeugen (50) eines Kurzzeitspektrums des Audiosignals, das eine Mehrzahl von Spektralwerten umfaßt;

Berechnen der psychoakustischen Maskierungsschwelle des Audiosignals unter Verwendung eines psychoakustischen Modells (58);

Quantisieren der Spektralwerte unter Berücksichtigung der psychoakustischen Maskierungsschwelle, so daß die durch die Quantisierung eingeführte Störenergie um einen vorbestimmten Betrag kleiner als die psychoakustische Maskierungsschwelle ist;

Bilden (56) eines Bitstroms, der Werte enthält, die den quantisierten Spektralwerten des Kurzzeitspektrums entsprechen.

13. Verfahren nach Anspruch 12, bei dem im Schritt des Bildens ferner eine Anzeige für den Wert (62) des vorbestimmten Betrags in den Bitstrom aufgenommen wird.

14. Vorrichtung zum Einbringen von Informationen in einen Datenstrom, der Daten über Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen, mit folgenden Merkmalen:

einer Einrichtung zum Verarbeiten (10, 16, 18) des Datenstroms, um die Spektralwerte des Kurzzeitspektrums des Audiosignals zu erhalten;

einer Einrichtung zum Beaufschlagen (32) der Informationen mit einer Spreizsequenz, um ein gespreiztes Informationssignal zu erhalten;

einer Einrichtung zum Erzeugen (34) einer Spektraldarstellung des gespreizten Informationssignals, um ein spektrales gespreiztes Informationssignal zu erhalten;

einer Einrichtung zum Ermitteln (40a; 40b; 40c; 40d) einer psychoakustisch maskierbaren Störenergie als Funktion der Frequenz für das Kurzzeitspektrum des Audiosignals, wobei die psychoakustisch maskierbare Stör-energie kleiner oder gleich der psychoakustischen Maskierungsschwelle des Kurzzeitspektrums ist;

einer Einrichtung zum Gewichten (36) des spektralen gespreizten Informationssignals unter Verwendung der ermittelten Störenergie, um ein gewichtetes Informationssignal zu erzeugen, bei dem die Energie der eingebrachten Informationen im wesentlichen gleich oder unterhalb der psychoakustischen Maskierungsschwelle liegt;

einer Einrichtung zum Summieren (38) des gewichteten Informationssignals mit den Spektralwerten des Kurzzeitspektrums des Audiosignals, um Summen-Spektralwerte zu erhalten, die das Kurzzeitspektrum des Audiosignals und die Informationen umfassen; und

einer Einrichtung zum Verarbeiten (22, 24, 26) der Summen-Spektralwerte, um einen verarbeiteten Datenstrom zu erhalten, der die Daten über die Spektralwerte des Kurzzeitspektrums des Audiosignals und die einzubringenden Informationen umfaßt.

15. Vorrichtung zum Codieren eines Audiosignals mit folgenden Merkmalen:

einer Einrichtung zum Erzeugen (50) eines Kurzzeitspektrums des Audiosignals, das eine Mehrzahl von Spektralwerten umfaßt;

einer Einrichtung zum Berechnen der psychoakustischen Maskierungsschwelle des Audiosignals unter Verwendung eines psychoakustischen Modells (58);

einer Einrichtung zum Quantisieren (52) der Spektralwerte unter Berücksichtigung der psychoakustischen Maskierungsschwelle, so daß die durch die Quantisierung eingeführte Störenergie gleich oder kleiner als die psychoakustische Maskierungsschwelle ist; und

einer Einrichtung zum Bilden (56) eines Bitstroms, der Werte enthält, die den quantisierten Spektralwerten des Kurzzeitspektrums entsprechen, und der darüberhinaus die berechnete psychoakustische Maskierungsschwelle (60) für das Kurzzeitspektrum des Audiosignals aufweist.

16. Vorrichtung zum Codieren eines Audiosignals, mit folgenden Merkmalen:

einer Einrichtung zum Erzeugen (50) eines Kurzzeitspektrums des Audiosignals, das eine Mehrzahl von Spektralwerten umfaßt;

einer Einrichtung zum Berechnen der psychoakustischen Maskierungsschwelle des Audiosignals unter Verwendung eines psychoakustischen Modells (58);

einer Einrichtung zum Quantisieren der Spektralwerte unter Berücksichtigung der psychoakustischen Maskie-

rungsschwelle, so daß die durch die Quantisierung eingeführte Störenergie um einen vorbestimmten Betrag kleiner als die psychoakustische Maskierungsschwelle ist;

einer Einrichtung zum Bilden (56) eines Bitstroms, der Werte enthält, die den quantisierten Spektralwerten des Kurzzeitspektrums entsprechen.

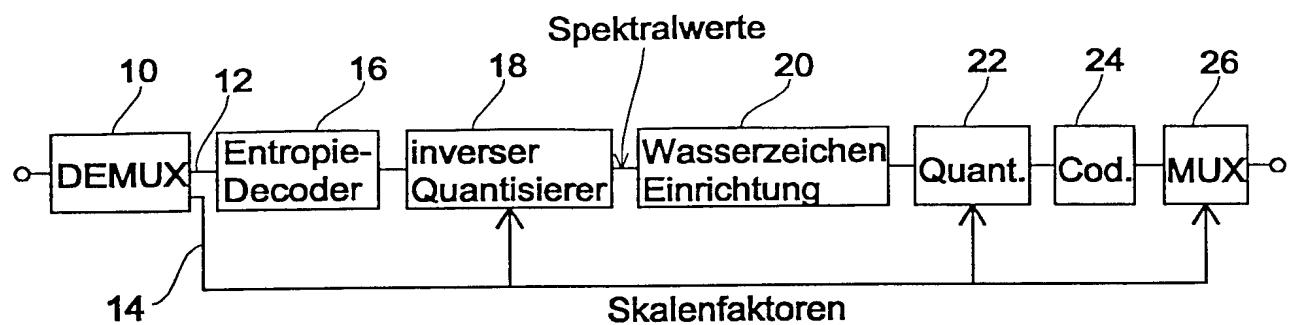


Fig. 1

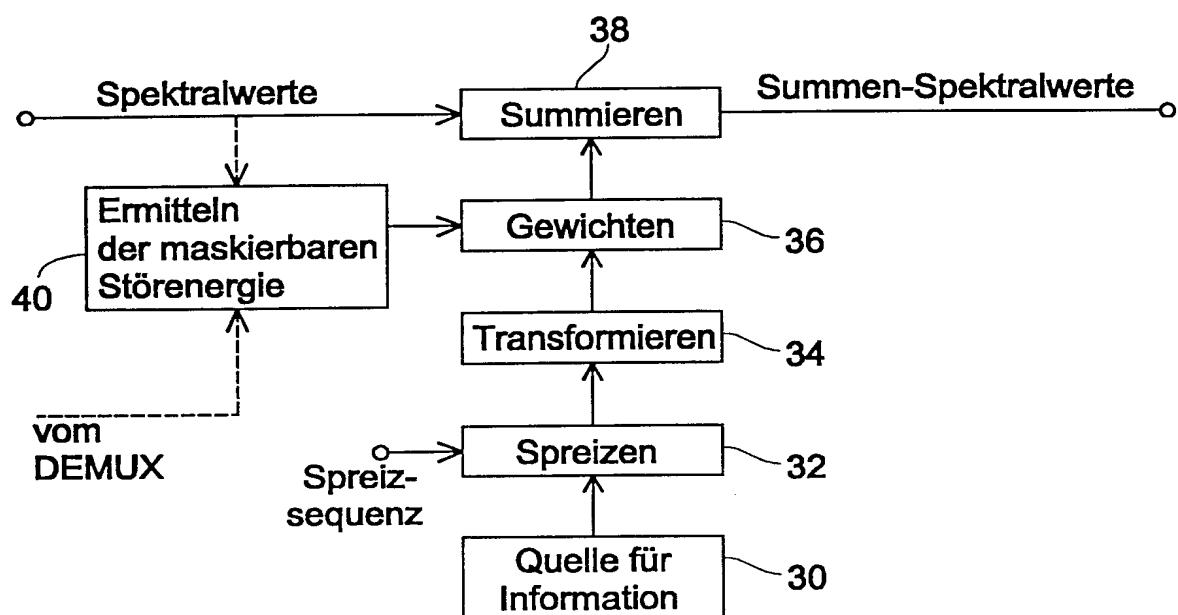


Fig. 2

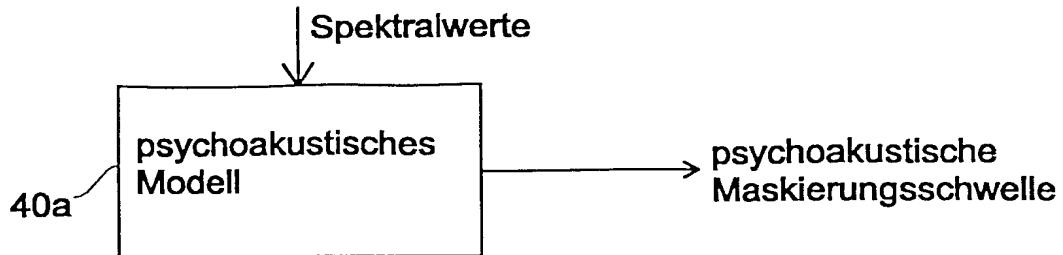


Fig. 3a

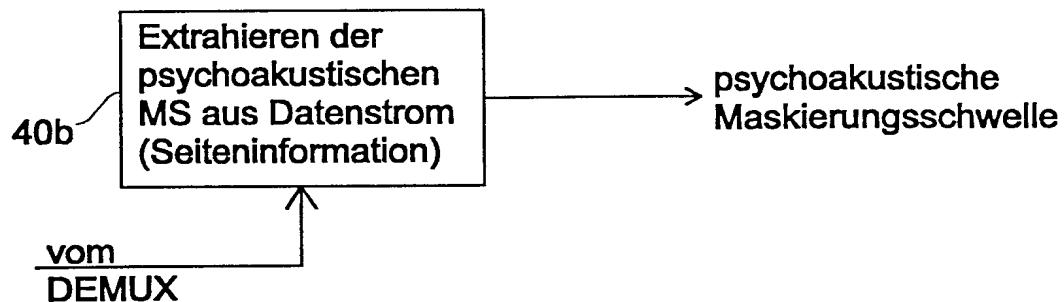


Fig. 3b

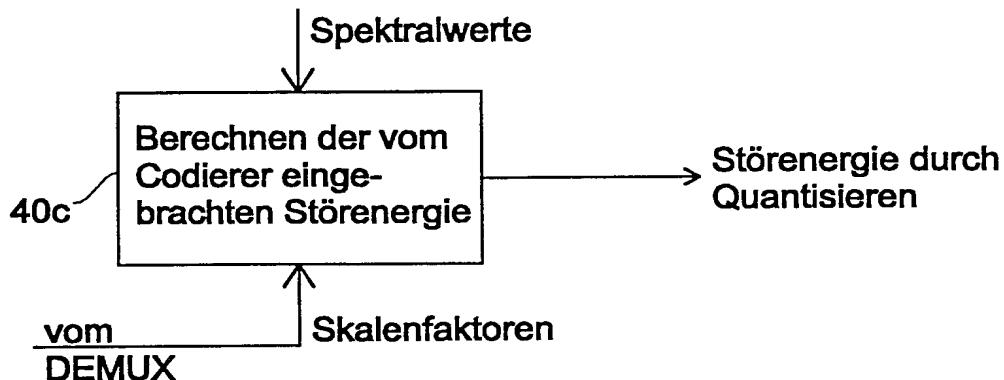


Fig. 3c

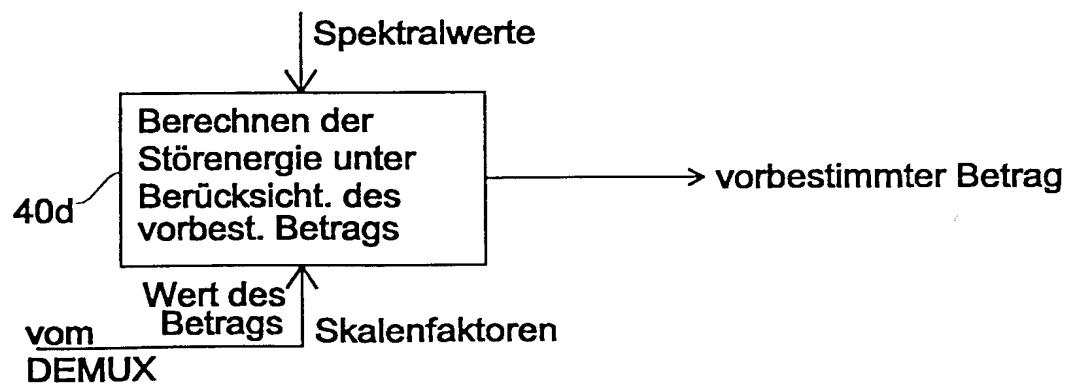


Fig. 3d

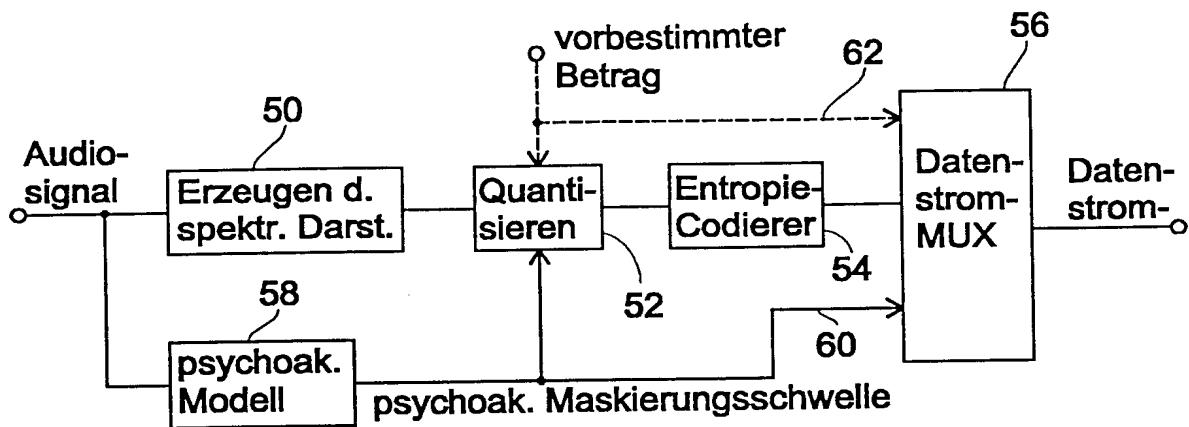


Fig. 4

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
12. April 2001 (12.04.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/26262 A3

(51) Internationale Patentklassifikation⁷: **H04H 1/00**,
H04B 1/66

ZUR FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. [DE/DE]; Leonrodstrasse 54,
80636 München (DE).

(21) Internationales Aktenzeichen: PCT/EP00/09771

(72) Erfinder; und

(22) Internationales Anmeldedatum:
5. Oktober 2000 (05.10.2000)

(75) Erfinder/Anmelder (nur für US): NEUBAUER, Christian [DE/DE]; Effeltricher Strasse 24, 90411 Nürnberg (DE). HERRE, Jürgen [DE/DE]; Am Eichengarten 11, 91054 Buckenhof (DE). BRANDENBURG, Karlheinz [DE/DE]; Haagstrasse 32, 91054 Erlangen (DE). ALLAMANCHE, Eric [DE/DE]; Sulzbacherstrasse 41, 90489 Nürnberg (DE).

(25) Einreichungssprache: Deutsch

(74) Anwälte: SCHOPPE, Fritz usw.; Schoppe, Zimmermann & Stöckeler, Postfach 71 08 67, 81458 München (DE).

(26) Veröffentlichungssprache: Deutsch

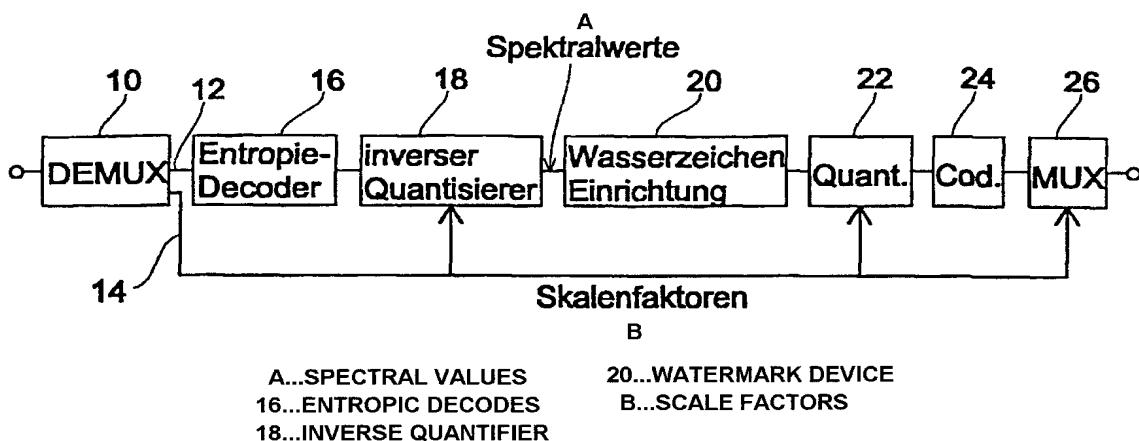
(81) Bestimmungsstaat (national): US.

(30) Angaben zur Priorität:
199 47 877.5 5. Oktober 1999 (05.10.1999) DE

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD AND DEVICE FOR INTRODUCING INFORMATION INTO A DATA STREAM AND A METHOD FOR
ENCODING AN AUDIO SIGNAL

(54) Bezeichnung: VERFAHREN UND VORRICHTUNG ZUM EINBRINGEN VON INFORMATIONEN IN EINEN DATEN-
STROM SOWIE VERFAHREN UND VORRICHTUNG ZUM CODIEREN EINES AUDIOSIGNALS



WO 01/26262 A3

(57) Abstract: The invention relates to a method for introducing information into a data stream containing data related to spectral values which represent a short-term spectrum of an audio signal. According to said method, the data stream is first processed (10, 16, 18), in order to obtain the spectral values for the short-term spectrum of the audio signal. In addition, the information to be introduced is subjected to a spread sequence (20) in order to obtain an expanded information signal which leads to the creation of a spectral representation of the expanded information signal (20). This representation is then weighted using a determined psychoacoustic noise energy which can be masked (20), in order to generate a weighted information signal, in which the energy level of the introduced information is substantially equal to or lies below the psychoacoustic masking threshold. The weighted information signal and the spectral values for the short-term spectrum are subsequently totalled (20) and then re-processed (22, 24, 26) in order to obtain a processed data stream which comprises both the audio information and the information to be introduced. In order for the information to be introduced without having to pass into the time domain, the block raster which underlies the short-term spectrum is not infringed, so that the introduction of a watermark does not lead to a tandem encoding effect.

[Fortsetzung auf der nächsten Seite]



(84) **Bestimmungsstaaten (regional):** europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht:
— mit internationalem Recherchenbericht

(88) **Veröffentlichungsdatum des internationalen Recherchenberichts:**

25. Oktober 2001

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

(57) **Zusammenfassung:** Ein erfundungsgemäßes Verfahren zum Einbringen von Informationen in einen Datenstrom, der Daten Über Spektralwerte aufweist, die ein Kurzzeitspektrum eines Audiosignals darstellen, führt zuerst eine Verarbeitung (10, 16, 18) des Datenstroms durch, um die Spektralwerte des Kurzzeitspektrums des Audiosignals zu erhalten. Außerdem werden die einzubringenden Informationen mit einer Spreizsequenz beaufschlagt (20), um ein gespreiztes Informationssignal zu erhalten, woraufhin eine spektrale Darstellung des gespreizten Informationssignals erzeugt wird (20), die dann mit einer ermittelten psychoakustisch maskierbaren Störenergie gewichtet wird (20), um ein gewichtetes Informationssignal zu erzeugen, bei dem die Energie der eingebrachten Informationen im Wesentlichen gleich oder unterhalb der psychoakustischen Maskierungsschwelle liegt. Das gewichtete Informationssignal und die Spektralwerte des Kurzzeitspektrums des Audiosignals werden dann summiert (20) und anschließend wieder verarbeitet (22, 24, 26), um einen verarbeiteten Datenstrom zu erhalten, der sowohl die Audioinformationen als auch die einzubringenden Informationen umfaßt. Dadurch, daß die einzubringenden Informationen in den Datenstrom eingebracht werden, ohne daß in den Zeitbereich Übergegangen werden muß, wird die Blockrasterung, die dem Kurzzeitspektrum zugrunde liegt, nicht angetastet, so daß das Einbringen eines Wasserzeichens zu keinen Tandem-Codiereffekten führt.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 00/09771

A. CLASSIFICATION OF SUBJECT MATTER
 IPC 7 H04H1/00 H04B1/66

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04H H04B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 687 191 A (LEE CHONG U ET AL) 11 November 1997 (1997-11-11) the whole document ---	1,2,4,6, 10,14 3,5
X	JOHNSTON J D: "TRANSFORM CODING OF AUDIO SIGNALS USING PERCEPTUAL NOISE CRITERIA" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, US, IEEE INC. NEW YORK, vol. 6, no. 2, 1 February 1988 (1988-02-01), pages 314-323, XP002003779 ISSN: 0733-8716 page 316, column 2, paragraph 2 -page 317, column 2, paragraph 3 page 319, column 2, paragraph C -page 320, column 2, paragraph G; figure 3 ---	11,15
Y	page 316, column 2, paragraph 2 -page 317, column 2, paragraph 3 page 319, column 2, paragraph C -page 320, column 2, paragraph G; figure 3 ---	3,5
		-/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

4 May 2001

10/05/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Pantelakis, P

INTERNATIONAL SEARCH REPORT

Intern. Application No

PCT/EP 00/09771

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KATE W R TH TEN ET AL: "DIGITAL AUDIO CARRYING EXTRA INFORMATION" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING. ICASSP, US, NEW YORK, IEEE, vol. CONF. 15, 3 April 1990 (1990-04-03), pages 1097-1100, XP000146962 the whole document ---	1,4,14
Y	DE 196 40 825 A (FRAUNHOFER GES FORSCHUNG) 11 September 1997 (1997-09-11) the whole document ---	1,4,14
X	EP 0 446 037 A (AMERICAN TELEPHONE & TELEGRAPH) 11 September 1991 (1991-09-11) page 1, line 22-33 page 5, line 30-39 page 6, line 11-22; claims 1,11,12,16; figures 1,5 ---	11,15
A	US 5 319 735 A (DERR ALAN G ET AL) 7 June 1994 (1994-06-07) the whole document -----	1,14

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern. Application No

PCT/EP 00/09771

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
US 5687191 A	11-11-1997	AU	1128797 A		27-06-1997
		EP	0873614 A		28-10-1998
		WO	9721293 A		12-06-1997
		US	5901178 A		04-05-1999
DE 19640825 A	11-09-1997	AT	184140 T		15-09-1999
		DE	59700389 D		07-10-1999
		WO	9733391 A		12-09-1997
		EP	0875107 A		04-11-1998
		DE	19640814 A		11-09-1997
EP 0446037 A	11-09-1991	CA	2037780 A,C		10-09-1991
		DE	69127842 D		13-11-1997
		DE	69127842 T		29-01-1998
		HK	1002743 A		11-09-1998
		JP	2774203 B		09-07-1998
		JP	4219799 A		10-08-1992
		JP	10133699 A		22-05-1998
		KR	185582 B		15-04-1999
		SG	44675 A		19-12-1997
US 5319735 A	07-06-1994	AT	190446 T		15-03-2000
		AU	676143 B		06-03-1997
		AU	3320793 A		19-07-1993
		DE	69230760 D		13-04-2000
		DE	69230760 T		23-11-2000
		EP	0617865 A		05-10-1994
		JP	7505984 T		29-06-1995
		WO	9312599 A		24-06-1993

INTERNATIONALER RECHERCHENBERICHT

Intern: als Aktenzeichen

PCT/EP 00/09771

A. KLASIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 7 H04H1/00 H04B1/66

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 7 H04H H04B

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, WPI Data, PAJ, INSPEC, COMPENDEX

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 5 687 191 A (LEE CHONG U ET AL) 11. November 1997 (1997-11-11)	1, 2, 4, 6, 10, 14
Y	das ganze Dokument ---	3, 5 .
X	JOHNSTON J D: "TRANSFORM CODING OF AUDIO SIGNALS USING PERCEPTUAL NOISE CRITERIA" IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, US, IEEE INC. NEW YORK, Bd. 6, Nr. 2, 1. Februar 1988 (1988-02-01), Seiten 314-323, XP002003779 ISSN: 0733-8716	11, 15
Y	Seite 316, Spalte 2, Absatz 2 -Seite 317, Spalte 2, Absatz 3 Seite 319, Spalte 2, Absatz C -Seite 320, Spalte 2, Absatz G; Abbildung 3 ---	3, 5
		-/-

Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

- *A* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- *E* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmelde datum veröffentlicht worden ist
- *L* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- *O* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- *P* Veröffentlichung, die vor dem internationalen Anmelde datum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

T Spätere Veröffentlichung, die nach dem internationalen Anmelde datum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

X Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

Y Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

*& Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

Absendedatum des internationalen Recherchenberichts

4. Mai 2001

10/05/2001

Name und Postanschrift der Internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Pantelakis, P

INTERNATIONALER RECHERCHENBERICHT

Intern: .ales Aktenzeichen

PCT/EP 00/09771

C.(Fortsetzung) ALS WESENTLICH ANGESEHEN UNTERLAGEN

Kategorie ^a	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	KATE W R TH TEN ET AL: "DIGITAL AUDIO CARRYING EXTRA INFORMATION" INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING. ICASSP, US, NEW YORK, IEEE, Bd. CONF. 15, 3. April 1990 (1990-04-03), Seiten 1097-1100, XP000146962 das ganze Dokument ---	1, 4, 14
Y	DE 196 40 825 A (FRAUNHOFER GES FORSCHUNG) 11. September 1997 (1997-09-11) das ganze Dokument ---	1, 4, 14
X	EP 0 446 037 A (AMERICAN TELEPHONE & TELEGRAPH) 11. September 1991 (1991-09-11) Seite 1, Zeile 22-33 Seite 5, Zeile 30-39 Seite 6, Zeile 11-22; Ansprüche 1,11,12,16; Abbildungen 1,5 ---	11, 15
A	US 5 319 735 A (DERR ALAN G ET AL) 7. Juni 1994 (1994-06-07) das ganze Dokument -----	1, 14

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internat...les Aktenzeichen

PCT/EP 00/09771

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
US 5687191 A	11-11-1997	AU	1128797 A	27-06-1997
		EP	0873614 A	28-10-1998
		WO	9721293 A	12-06-1997
		US	5901178 A	04-05-1999
DE 19640825 A	11-09-1997	AT	184140 T	15-09-1999
		DE	59700389 D	07-10-1999
		WO	9733391 A	12-09-1997
		EP	0875107 A	04-11-1998
		DE	19640814 A	11-09-1997
EP 0446037 A	11-09-1991	CA	2037780 A,C	10-09-1991
		DE	69127842 D	13-11-1997
		DE	69127842 T	29-01-1998
		HK	1002743 A	11-09-1998
		JP	2774203 B	09-07-1998
		JP	4219799 A	10-08-1992
		JP	10133699 A	22-05-1998
		KR	185582 B	15-04-1999
		SG	44675 A	19-12-1997
US 5319735 A	07-06-1994	AT	190446 T	15-03-2000
		AU	676143 B	06-03-1997
		AU	3320793 A	19-07-1993
		DE	69230760 D	13-04-2000
		DE	69230760 T	23-11-2000
		EP	0617865 A	05-10-1994
		JP	7505984 T	29-06-1995
		WO	9312599 A	24-06-1993